AD-A062 494

GEORGIA INST OF TECH ATLANTA SCHOOL OF INFORMATION A--ETC F/6 12/1
ON SMALL UNIVERSAL DATA STRUCTURES AND RELATED COMBINATORIAL PR--ETC(U)
AUG 78 R A DEMILLO'S EISENSTAT', R J LIPTON DAAG29-76-G-0338

UNCLASSIFIED

END
AUGUS

END
AUGUS

3 - 79
DOC

END
AUGUS

ACCURATE ATLANTA SCHOOL OF INFORMATION A--ETC F/6 12/1
DAAG29-76-G-0338

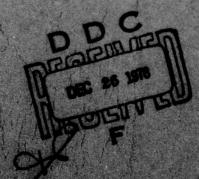
ARO-14690.5-EL

NL

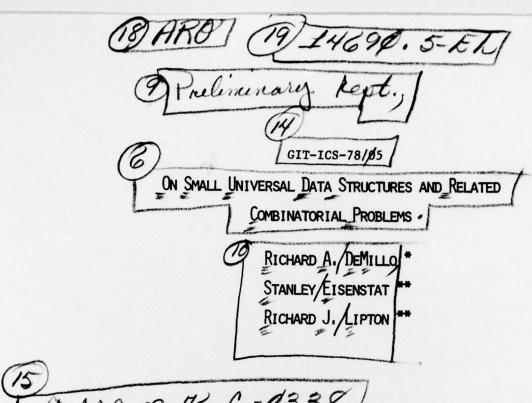
EVE ARO 14690, 5-82

ADA062494

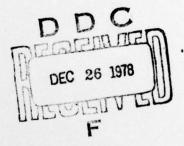


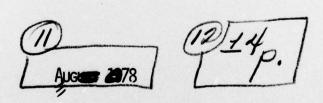


DOC FILE COPY



DAAG-29-76-6-03381





This document has been approved for public release and sale; its distribution is unlimited.

- Georgia Institute of Technology
- Yale University

440 944

ON SMALL UNIVERSAL DATA STRUCTURES AND RELATED COMBINATORIAL PROBLEMS † (Preliminary Report)

Richard DeMillo

School of Information and Computer Science Georgia Institute of Technology Atlanta, GA 30332

> Stanley Eisenstat Richard J. Lipton

Department of Computer Science Yale University New Haven, CT 06520

[†] Work supported in part by U.S. Army Research Office, Grant No. DAAG29-76-G-0338, and by the Mathematics Research Center, University of Wisconsin-Madison.

INTRODUCTION

One of the most significant changes in theoretical computer science has been the recent infusion of the methods and problems from combinatorial analysis. Among the most powerful combinatorial theorems which have been imported to computer science are those of extremal graph theory [1]: in extremal graph theory, one is interested in the largest (or in complementary problems, the smallest) graph which avoids (or contains) a given structure. Purely combinatorial results (which have significance, e.g., for the design of circuit boards) have been obtained by Chung and Graham [2] and by Chung, Graham, and Pippenger [3]. In this paper, we extend this theory to encompass results concerning data structures.

As motivation for the results to the described, note that many of the large data structures manipulated by the programs described in [4,5] have two characteristics

- (i) they are sequentially accessed, and
- (ii) many distinct structures convolve in the same physical memory.

For applications of this sort, it would obviously be desirable to have available a universal data structure in which all data structures from a given class may gracefully reside. In view of (i), by "graceful" we mean that the sequential accessing characteristics of the embedded data structures are not too drastically altered. Let us measure such alterations by the dilation of logical adjacencies [6,7] needed to embed all structures from a given class into a universal structure; this is then a complementary extremal graph theory problem: what is the size (number of edges) of the smallest universal graph for a given dilation factor

The main results contained in this paper address such problems from a number of points of view.

- (1) We give several asymptotically optimal universal data structures for graphs of n vertices when average dilation [7] is used as a measure.
- (2) We discuss a universal data structure for graphs of n vertices where worst-case dilation is used as a measure [6].
- (3) We consider variations of the average dilation measure which gives favorable comparisons between data structures studied in [6,7].
- (4) We consider the kinds of "sharing" that can take place between "almost linear" and "almost complete tree-like" structures.
- (5) Finally, we propose a data structure embedding model which recovers some aspects of random accessing of data items, and prove a space-time tradeoff which seems to indicate that no savings is possible in RAM models which assess accessings costs uniformly [8].

PRELIMINARIES

A graph, G, is defined by its vertices, V(G), and edges, $E(G) \subseteq V(G) \times V(G).$ Edges are assumed to be undirected: a pair of vertices x,y are connected if either $(x,y) \in E(G)$ or $(y,x) \in E(G)$. A path between x_0, x_n is said to be of length n. The distance metric $d_G(x_0, x_n)$ is defined to be n if there is no shorter path than x_0, \ldots, x_n .

A graph represents a data structure in the obvious way: vertices represent nodes or records and connectedness models logical adjacency. The following relations and their significance for data structures can be found in [6,7]. Let G, G* be graphs. We say that G is T-worst case embeddable in G* ($G \subset G$) if there is a one-one $\Phi: V(G) \to V(G^*)$ such that $(x,y) \in E(G)$ implies

$$d_{G^*}(\Phi(x), \Phi(y)) \leq T.$$
 (1)

Similarly, G is A-average case embeddable in G* (G $\leq \frac{\text{avg}}{A}$ G*) if there is a one-one Φ as above such that

$$\sum_{\mathbf{x},\mathbf{y}} d_{\mathbf{G}^{*}}(\Phi(\mathbf{x}), \Phi(\mathbf{y})) \leq A \cdot |E(G)|.$$
connected

In [4,5], comparisons between several natural classes of graphs give asymptotic bounds on T, A in (1), (2) as functions of |V(G)|. Shortly after the announcement of the results of [6], R. M. Karp suggested to us the following class of problems connected with extremal graph theory: what are the characteristics of \leq_T - universal data structures; i.e., those structures which T-worst case embed all graphs in a given class. This paper grew out of considering these problems.

UNIVERSAL GRAPHS

Let ζ^n be a given class of graphs G, |V(G)| = n. Let us ask about a data structure which is \leq_T or $\leq \frac{avg}{A}$ universal for ζ^n . In particular, let us define

$$w(\zeta^n, T) = \underline{\min} \{|E(G)|: G^n \in \zeta^n, G^n \leq_T G\}$$
 (3)

and

$$a(\zeta^n, A) = \underline{\min} \{|E(G)|: G^n \in \zeta^n, G^n \leq \frac{avg}{A} G\}.$$

For T = 1, (3) becomes the complementary extremal graph problem studied in [2,3].

By an *n-tree* G, we mean a connected acyclic graph G, with |V(G)| = n. It is also convenient to think of trees as rooted in the following sense: accompanying G, there is an ancestor-descendent relation that assigns direct ancestors and direct descendents to vertices in the obvious way so that a vertex with no ancestors can be designated as the root of the tree.

(Obviously this choice is not going to be unique, but we assume that G is not characterized until such a choice is made). A d-ary n-tree is an n-tree in which each vertex has at most d direct descendents. We denote, respectively, the classes of n-trees and d-ary n-trees by Γ^n and Γ^{n*}_d .

By [2] it is known that $\frac{1}{2}n \log n < w(\Gamma^n, 1) < n^{1+k(n)}, k(n) = [\log \log n]^{-1}$.

The upper bound was improved in [3] to

$$w(\Gamma^n, 1) = 0(n \log n[\log \log n]^2)$$

The bounds on $a(\Gamma^n, 1)$ are apparently not elsewhere considered.

Superficially, at least, all interest in further characterization of

(3) is destroyed by the following obvious

Theorem. For $T \geq 2$

$$w(\Gamma^n, T) = n$$

Of course, in (3), the "target" graph G may have unbounded degree. Therefore, it is natural to consider $w(\zeta^n, T, S)$ and $a(\zeta^n, T, S)$ where in both cases the target graph G is restricted to be in the set S. Note that now the theorem just cited is no longer obviously true.

^{*} Thus Γ_2^n = binary trees on n vertices.

⁺ In the sequel, we use log x for log₂x and lnx for log_ex.

The best that is known is the upper bound of [3] (S = all cubic graphs)

$$w(\Gamma^n, 1, S) \le \frac{2\sqrt{2}}{n} \exp(\log^2 n/2 \log 2)$$
. (4)

It is not obvious that when (i) "targets" are restricted to binary trees and (ii) $w(\Gamma_2^n, T, \Gamma_2^n)$ is considered, that it is possible to do any better than the union of all trees in Γ_2^n , giving a structure of size $4^n/_{2\pi n}$. But, we have the following

Theorem. For each T \geq 1, there is a binary tree H, such that G \geq_T H for all G $\in \Gamma_2^n$, and

$$\ln |E(G)| \leq \frac{\ln^2 n}{\ln 4}$$
;

or in other words

$$w(\Gamma_2^n, T, \Gamma_2^n) = \exp \frac{1}{\ln 4} (\ln n)^2 + O((\ln n)^2)$$
.

A key step in the proof of this theorem hinges on the solution to the fascinating "almost linear" recurrence

$$u_n = u_{n-1} + u_{n-1}$$
, (5)

first considered by Knuth [9]. This also establishes a connection between the theorem and ineq. (4): u_n is also the number of partitions of 2n of the form $\Sigma \alpha_i^2$, $\alpha_i^2 = 0$, 1. Knuth [9] bounds the partition function

$$P(m) = \frac{1}{4\sqrt{3m}} \exp{(\pi \frac{2}{3} m)}$$
.

There are two possibilities for improving the bounds in $w(\Gamma_2^n, T, \Gamma_2^n)$. The first possibility is to introduce circuits to the target graph of the previous theorem, but this does not appear to give an asymptotically better bound than (4). The second possibility is to prove that balanced trees and unbalanced trees are \leq_T - equivalent. This seems unlikely since combining such a result with the proof method of the previous theorem gives a polynomial sized universal tree. However, in trying to improve the bounds on $w(\Gamma_2^n, T, \Gamma_2^n)$ it may be desirable to ignore irregular trees, letting only very balanced or very unbalanced trees reside in the same universal data structure.

In any case, it seems unlikely that polynomial structures are possible. We are, however, far from proving this; indeed, the best known lower bound is the following

Theorem. For all n > N

$$w(\Gamma_2^n, T, \Gamma_2^n) > c(T) n \log n$$

where c(T) > 0 is a constant for fixed $T \ge 1$.

Certain other subcases are also of interest. Erdős, Chung, and Graham[†], consider w(S,1) and obtain

$$w(S,1) \leq \frac{4}{11} n^2$$
.

The following theorem is an improvement, but is surely not the best possible bound.

Theorem

$$w(S,1) \leq \frac{2}{9} n^2$$

[†] Private Communication.

A non-trivial lower bound would clearly be desirable. Another class of interest are graphs of high genus. $^{++}$ We conjecture that for graphs of fixed genus γ , it is possible to do better than the naive $\binom{n}{2}$ bound obtained by embedding in the complete graph.

Our next series of results show impressive improvements by passing to average dilations. We now get optimal constructions, even in a variety of limited settings.

We have, for instance, the

Theorem. For $\alpha > 0$,

$$a(\Gamma_2^n, \frac{1}{\alpha}, S) = 0(n^{\log(2+\alpha)})$$
.

Since there is a linear lower bound on $a(\cdot,\cdot,\cdot)$, this construction is optimal. By a slight modification of the construction, this gives $a(\Gamma_2^n, A, S) = O(n)$, for all $A \ge 1$, but this result may be superceeded by the following

Theorem. For each $A \ge 1$, there is a binary tree H, such that

$$G \leq \underset{A}{\text{avg}} H$$

for all $G \in \Gamma_2^n$, and

$$|E(G)| = O(n)$$
;

or, in other words

$$a(\Gamma_2^n, A, \Gamma_2^n) = 0(n)$$
.

^{††} A graph is of genus γ if it can be embedded in a sphere with γ handles [10].

These results are related to the ability to "cut" graphs in advantageous ways. For example, a generalization of the planar separator theorem [11] to graphs of high genus, obtained by Lipton and Tarjan, gives us the following

Theorem. Let L_{γ}^{n} be the class of graphs G with genus γ and |V(G)| = n. Then, for all n > N,

$$a(L_{\gamma}^{n}, A, \Gamma_{2}^{n}) \leq c(A) \cdot n$$
,

where c(A) does not depend on n.

EXTENDED MODEL

In comparing classes of data structures (see, e.g., [6,7], the measures of "efficiency" have implicitly assumed that only sequential accessing is important. Thus, when in [6], we bound the *efficiency*, T, of an embedding of $n \times n$ array into binary trees by

$$T \ge c \log n$$
,

the function T(n) captures the dilation factor in an embedding. We now describe a generalization of this concept which recovers a certain kind of random accessing. Since the precise definitions are quite complex, we will settle for a less exact -- but more picturesque -- rendering. Let us assume that we have in front of us an illustration of a graph G, and also a number of friends who agree to lend us their forefingers for use in tracing the paths of the graph. Our friends oblige us as follows: We may start traversing at any vertex already visited. The traversal rule is, then, that we must either traverse graph edges or "jump" to a vertex pointed to

by a friend. The *time* required to traverse a sequence of vertices is then simply the number of applications of traversal rules. Notice that the result of a traversal is not necessarily a path of G. The connection between fingers and random accessing is that traversals requiring k-fingers also require k-"addresses" for the vertices pointed to.

We then say that $G \leq_{k,T} G^*$ if there is a one-one $\Phi: V(G) \to V(G^*)$, so that for every $x,y \in V(G)$ with $d_G(x,y) = m$, there is a k-finger traversal from $\Phi(x) = x^*$ to $\Phi(y) = y^*$ with time at most Δ , and $\Delta \leq Td_{G^*}(x^*, y^*)$. We have the following

Theorem. If G is the $n \times n$ array [7], H is a binary tree and

$$G_n \leq k, T(n)$$
 H,

then

$$k + T(n) > c \log n$$
,

where c is a constant independent of n.

OTHER TYPES OF AVERAGE EMBEDDING

The relation $\leq \frac{\text{avg}}{A}$ may be thought of as averaging - with relative frequencies uniformly distributed to the edges E(G) - over the edges of G. We now make a more global definition which may be used to recover our intuitions about path lengths in binary trees [7]. We will essentially average our shortest paths:

 $G \leq \frac{\text{paths}}{A}$ G* if there is an embedding $\Phi: V(G) \to V(G^*)$ such that

$$\sum_{\substack{d \\ \Phi(x), \Phi(y)}} (\Phi(A, \Phi(y)) \leq A \cdot \sum_{x,y} d_{G}(x,y) .$$

We then have the following

For each $n \ge 0$, let A be the least real number such that

$$G_n \leq \frac{\text{paths}}{A} H$$
,

for a binary tree H. Then

$$\lim_{n\to\infty}A_n=0$$

Thus, we see that if the average embedding is required to work well on all shortest paths, then the embedding cost goes to zero. In a sense, then < avg "charges" more heavily than < paths for any bottlenecks.

REFERENCES

- [1] P. Erdös and J. Spencer. Probabilistic Methods in Combinatorics. Academic Press, 1974.
- [2] F. R. K. Chung and R. L. Graham. On Graphs Which Contain Small Trees. To appear in JCT.
- [3] R. R. K. Chung, R. L. Graham, and N. Pippenger. On Graphs Which Contain All Small Trees, II. To appear in JCT.
- [4] M. Minsky. Semantic Information Processing. MIT Press, 1969.
- [5] L. Uhr. Pattern Recognition Learning and Thought. Prentice-Hall, 1973.
- [6] R. J. Lipton, S. C. Eisenstat, and R. A. DeMillo. Space and Time Hierarchies for Classes of Control and Data Structures. JACM 23(4): 720-732, October 1976.
- [7] R. A. DeMillo, S. C. Eisenstat, and R. J. Lipton. Preserving Average Proximity In Arrays. CACM, Vol. 21(3):228-231, March 1978.
- [8] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. The Design and Analysis of Computer Algorithms. Addison-Wesley, 1976.
- [9] D. E. Knuth. An Almost Linear Recurrence. The Fibonacci Quarterly, Vol. 4(1):117-128, February 1966.
- [10] F. Harary. Graph Theory. Addison-Wesley, 1972.
 [11] R. Lipton and R. Tarjan. A Planar Separator Theorem. Stanford Research Report CS-77-627, October 1977.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Date Entered)

	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2. GOVT ACCESSION NO	. 3. RECIPIENT'S CATALOG NUMBER
877-TCS-77/05	
4. TITLE (and Subtitio)	5. TYPE OF REPORT & PERIOD COVERED
On Small Universal Data Structures and	
Related Combinatorial Problems	TECHNICAL
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(a)	8. CONTRACT OR GRANT NUMBER(a)
Richard A DeMillo, Stanley Eisenstat,	DAAG29-76G-0338
Richard J. Lipton	Dialo25 700 0550
Richard 3. Elpton	
PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
School of Information and Computer Science	AREA & WORK UNIT NUMBERS
Georgia Institute of Technology	
Atlanta, Georgia 30332	
1. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
	August, 1978
	13. NUMBER OF PAGES 10 + ii
4. MONITORING AGENCY NAME & ADDRESS(II different from Controlling Office)	15. SECURITY CLASS. (of this report)
US Army Research Office	13. SECONITY CEASS. (of time report)
PO Box 12211	Unclassified
Research Triangle Park, N.C. 27709	16. DECLASSIFICATION/DOWNGRADING
, , , , , , , , , , , , , , , , , , , ,	SCHEDULE
Approved for Published Releases; Distribution Un	nlimited
7. DISTRIBUTION STATEMENT (of the abetract entered in Block 20, If different in Block 20, If dif	om Report)
7. DISTRIBUTION STATEMENT (of the abetract entered in Block 20, If different in Block 20, If dif	om Report) ed as an official Department
7. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, If different in Block 20, If dif	om Report) ed as an official Department
7. DISTRIBUTION STATEMENT (of the abetract entered in Block 20, If different in Block 20, If dif	om Report) ed as an official Department
7. DISTRIBUTION STATEMENT (of the abetract entered in Block 20, if different in Block 20, if dif	ed as an official Department er authorized document.
7. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different in Block 20, if dif	ed as an official Department er authorized document.
7. DISTRIBUTION STATEMENT (of the abetract entered in Block 20, if different in Block 20, if dif	ed as an official Department er authorized document.
7. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different in Block 20, if dif	ed as an official Department er authorized document.
7. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different in Block 20, if dif	ed as an official Department er authorized document.
The findings of this report are not to be constru of the Army position, unless so designated by oth KEY WORDS (Continue on reverse side if necessary and identify by block number Graph embedding, data structures, universal graph access.	ed as an official Department er authorized document.
7. DISTRIBUTION STATEMENT (of the abetract entered in Block 20, if different in Block 20, if dif	ed as an official Department er authorized document. s, binary trees, random
7. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different in a supplementary notes The findings of this report are not to be construed to the Army position, unless so designated by other words (Continue on reverse side if necessary and identify by block number Graph embedding, data structures, universal graph access. ABSTRACT (Continue on reverse side if necessary and identify by block number access. ABSTRACT (Continue on reverse side if necessary and identify by block number access.	ed as an official Department er authorized document. s, binary trees, random raph G such that every average case) loss of al graph can be interpreted
8. SUPPLEMENTARY NOTES The findings of this report are not to be construed to the Army position, unless so designated by othe Army position, unless so designated by othe Graph embedding, data structures, universal graph access. ABSTRACT (Continue on reverse side if necessary and identify by block number, a universal graph for a class of graphs T is a graph in T can be embedded in G with worst-case (proximity bounded by a fixed constant. A universal	ed as an official Department er authorized document. s, binary trees, random raph G such that every average case) loss of al graph can be interpreted